Spatial Audio Engineering in a Virtual Reality Environment

Dennis Jordan, Fabian Müller, Colin Drude, Sascha Reinhold,
Vanessa Schomakers, Michael Teistler

Faculty of Information and Communication, Flensburg University of Applied Sciences

Abstract

The topic of audio engineering in virtual reality is relatively unexplored. We introduce an approach for manipulating sound and music in space, oriented towards amateurs using a VR head-mounted display and hand-held motion controls. An exploratory usability test shows that in our application spatial audio is well received. However, in the future we would like to improve the interaction design with regard to the recording functionality. We assume that we are at the starting point of a broader research interest in spatial audio engineering in VR due to the increasingly affordable hardware.

1 Introduction

Audio engineers manipulate the recordings of artists by using analog or digital mixing tables and sequencers. Over the course of a few years virtual reality (VR) has become a more common form of interactive visualization (Statista GmbH 2016). VR allows users to experience virtual three-dimensional space via stereoscopic representation and head tracking. In this study a VR head-mounted display (HMD) and motion controls are used for an audio engineering application, oriented towards amateurs with focus on spatial sound. By using motion controls, inexperienced users should be able to easily change the location of sound sources in a virtual environment. The process of editing is supported by visual and audible feedback.

While there are many articles examining audio in augmented and virtual reality (e.g., Härmä et al. 2004; Lentz et al. 2007; Zhao et al. 2016), the topic of audio engineering in VR is relatively unexplored. The idea to use gesture based motion control input for audio engineering was already explored (Drossos et al. 2013; Kumbeiz & Hlavacs 2014). However, in these cases the users were not provided with the precise controls that we intended to implement.

Veröffentlicht durch die Gesellschaft für Informatik e.V. 2016 in
S. Franken, U. Schroeder, T. Kuhlen (Hrsg.):
Mensch und Computer 2016 – Kurzbeiträge, 4. - 7. September 2016, Aachen.
Copyright © 2016 bei den Autoren.
http://dx.doi.org/10.18420/muc2016-mci-0217

2 Materials and Methods

2.1 Utilized Hardware and Interaction Techniques

The prototype was built with *Unity 3D* 5.34 (Unity Technologies 2016) and uses the *Oculus Rift DK2* HMD (Oculus VR LLC. 2016) as well as *Razer Hydra* (Sixense Entertainment Inc. 2016) game controllers (Figure 1). The game controllers represent the user's hands and are used to change the position of the sound sources through direct manipulation (cf. Shneiderman 1983). The sound sources are visually represented as cubes with an image of the instrument that is played. Imitating the human process of grabbing and placing things, users press the trigger and move the controllers to move the cubes. The cube shape was chosen as an analogy for building blocks, utilizing something familiar for inexperienced VR users. By pressing any button on the top of the controller the hand changes visually from a grabbing gesture (Figure 2 - left) to a pointing gesture (Figure 2 - right). The user can now start and stop sounds by aiming the virtual index finger at the desired sound source and pulling the trigger. To assist the user with the aiming, a virtual laser pointer beam is shown, starting from the tip of the index finger. Pressing any button on the top again toggles the gesture back to its original state. We visualized those features with different gestures, so that inexperienced users have a clear distinction between those two control modes. With the HMD being a closed environment, the user cannot see the controllers.

Additionally, we used stereo headphones, enabling the user to receive audible feedback based on head movement. To achieve that, we utilized the internal audio handling of the Unity game engine. With headphones the user can determine the direction the sounds are coming from, which is not possible with a stationary stereo speaker setup.

Figure 1: Application setup

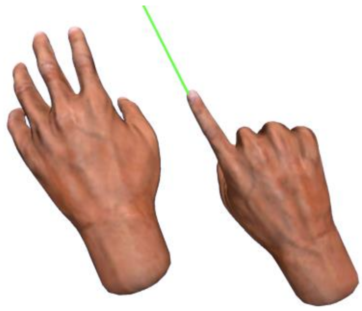

Figure 2: Gestures of virtual hands.
Left: Grabbing pose. Right: Pointing pose.

With the prototype the user is able to compose an audio track based on predefined sound sources. Upon application startup the user is located in a virtual room with different interactive sound sources in front of him/her as well as a virtual record button and a virtual play button. These buttons resemble those of a tape recorder. The user can interact with the sound sources using the game controllers. The interaction comprises repositioning, playing and stopping audio. Once the user is pleased with the position of the sound sources, pressing the virtual record button by grabbing it starts the recording. While recording, the user can start and stop sound sources as well as move them around. Each action is saved until the recording is stopped by pressing the record button. The recording can be replayed by pressing the green play button. An example screenshot is shown in figure 3.

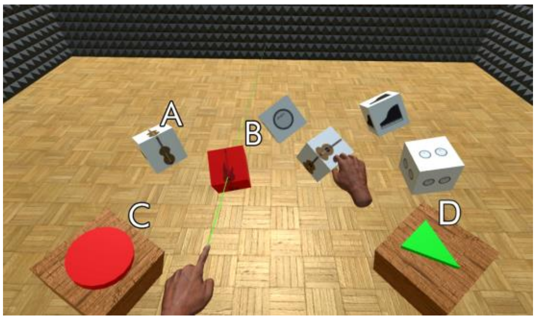

Figure 3: Screenshot of the prototype.
A: Representation of inactive sound, B: Representation of active sound,
C: Record button, D: Play button

2.2 Exploratory Usability Test

For an exploratory usability test, we used a combination of a think-aloud protocol and an observation, followed by inquiring further feedback from the participants to explore how an amateur sound engineer would use and assess our approach. The prototype was tested by a group of six users, who were inexperienced in audio engineering applications. The participants were given an introduction to the basic functionality of the game controllers and the idea that the software is intended for recording spatial audio. Afterwards the participants tried the software out for a short amount of time in order to get familiar with the controls. Then the recording action was explained to the users and they were asked to record an audio track with the given sound sources.

3 Results

We found that all participants quickly understood that the sounds were mapped onto the cubes. They had no problem with grabbing and placing the sound sources. At first, one participant had difficulties using the virtual hands, because the positions of the game controllers are not exactly mapped to those of the virtual hands. Most of the participants had no problem understanding how to play and pause the sound sources, but were not able to aim correctly. One participant had difficulties to switch the control mode. Before the introduction to the recording was given, the participants did not understand the record and play buttons in the application. While recording, it was hard to activate an audio sound source in time to synchronize it with the other sound sources. Only one of the participants moved the sound sources while recording.

4 Discussion

We assume that changing the position of sound sources was quickly understood because of the direct manipulation of the sound sources. Additionally, the depth perception of the HMD probably helped to locate the sound sources. Positive feedback was given on the overall spatial audio experience. The participants were pleased how well it worked and tested the sound positioning in many different configurations. Negative feedback relates to the playing and stopping of the audio as well as the recording functionality. The users had problems aiming at the sound sources and therefore using them correctly. We assume that there should be additional visual feedback when the virtual laser pointer beam hits a sound source. Without further instructions the recording function was not fully understood, which lead to confusion. Most of the participants did not move the sound sources while recording, potentially because the users did not expect this to be possible. In future work we intend to improve and evaluate our recording concept with regard to the overall user experience.

References

Drossos, K., Floros, A., & Koukoudis, K. (2013). Gestural User Interface for Audio Multitrack Real-time Stereo Mixing. *AM'13 Proceedings of the 8th Audio Mostly Conference*.

Härmä, A., Jakka, J., Tikander, M., Karjalainen, M., Lokki, T., Hiipakka, J., & Lorho, G. (2004). *Augmented Reality Audio for Mobile and Wearable Appliances.* AES.

Kumbeiz, W., & Hlavacs, H. (2014). Bass Impact DJ Studio – A Music Mixing Software with Gesture Controls. *Intelligent Technologies for Interactive Entertainment*, pp. 132 - 135.

Lentz, T., Schroder, D., Vorlander, M., & Assenmacher, I. (2007). Virtual Reality System with Integrated Sound Field Simulation and Reproduction. *EURASIP Journal on Advances in Signal Processing*.

Oculus VR LLC. (2016). *DK2 Development Kit 2*. Retrieved June 02, 2016, from oculus.com: https://www.oculus.com/en-us/dk2/

Shneiderman, B. (1983). Direct Manipulation: A step beyond programming languages. *Human Factors in Interactiue Computer Systems*.

Sixense Entertainment Inc. (2016). *Razer Hydra*. Retrieved June 2016, 2016, from sixense.com: http://sixense.com/razerhydra

Statista GmbH. (2016). *Statista GmbH*. Retrieved June 02, 2016, from Statista.com: http://www.statista.com/statistics/509154/head-mounted-displays-worldwide-shipments

Unity Technologies. (2016). *Unity Technologies*. Retrieved June 02, 2016, from unity3d.com: https://unity3d.com/unity

Zhao, S., Ahmed, S., Liang, Y., Rupnow, K., Chen, D., & Jones, D. L. (2016). *A Real-Time 3D Sound Localization System with Miniature Microphone Array for Virtual Reality*. IEEE.